REMARKS

Claims 1-44 remain in this application. Claim 34 has been amended. Claims 1-44 are pending.

The Specification has been amended to reflect the current status of referenced co-pending patent applications. Claim 34 has been amended to correct a typographical error. No claim has been amended in response to the 35 U.S.C. 103(a) rejection.

Claims 1-44 stand rejected under 35 U.S.C. 103(a) as being obvious over U.S. Patent No. 6,510,406, to Marchisio ("Marchisio"), in view of U.S. Patent No. 6,701,305, to Holt et al. ("Holt"). Applicant traverses the rejection. To establish a *prima facie* case of obviousness: (1) there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or combine the reference teachings; (2) there must be a reasonable expectation of success; and (3) the combined references must teach or suggest all the claim limitations. MPEP § 2143. A *prima facie* case of obviousness has not been shown.

Marchisio discloses an information retrieval system that allows a wide margin of uncertainty in the initial choice of keywords in a query (Abstract). The system computes a constrained measure of the similarity between a query vector and all documents in a term-document matrix (Col. 6, lines 35-39). The system parses electronic information files containing text, which may include recognizing acronyms, recording word positions and extracting word roots (Col. 6, lines 35-43). The parsing may further include generating a number of concept identification numbers corresponding to respective terms to be associated with the rows of a term-document matrix and the counting of individual terms in each of the files (Col. 6, lines 47-54). The system generates the term-document matrix based on the contents of the files parsed and the value of each cell indicates the number of occurrences of the respective term within one of the files, or, alternatively, the value of the cell may reflect the presence or absence of the respective term (Col. 6, lines 55-65). The system receives a user query from a user, consisting of a list of keywords or phrases that are parsed to generate a query

vector (Col. 5, lines 8-14; Col. 7, lines 27-39). The similarity between the query and document projections are measured as a constrained optimization problem in a linear transform space, which maximizes the stability of a solution at a given level of misfit (Col. 5, lines 14-17). The system may decompose the term-document matrix in terms of orthogonal basis functions and each basis encodes groups of conceptually related keywords that are arranged in order of decreasing statistical relevance to the query (Abstract).

Holt discloses retrieving information from a text data collection that comprises a plurality of documents, which each consist of a number of terms (Abstract; Col. 6, lines 12-16). The text data collection is represented by a term-by-document matrix having a plurality of entries with each entry representing the frequency of occurrence of a term in a respective document (Col. 6, lines 16-19). An orthogonal basis for a lower dimensional subspace is generally obtained from the term-by-document matrix as part of document indexing (Col. 6, lines 19-22). A representation of at least a portion of the original matrix is projected into a lower dimensional subspace and those portions of the subspace representation that relate to terms of a query are weighted (Col. 5, lines 62-67). A plurality of documents are scored with respect to the query based at least partially upon the weighted portion of the subspace representation (Col. 6, lines 31-33). Documents can then be identified based upon ranking the scores of the documents with respect to the query (Col. 6, lines 34-36).

In contrast, Claim 1 defines a system for analyzing unstructured documents for conceptual relationships. Claim 1 recites a histogram module determining a frequency of occurrences of concepts in a set of unstructured documents, each concept representing an element occurring in one or more of the unstructured documents. Claim 1 further recites a selection module selecting a subset of concepts out of the frequency of occurrences, grouping one or more concepts from the concepts subset, and assigning weights to one or more clusters of concepts for each group of concepts. Claim 1 further recites a best fit module calculating a best fit approximation for each document indexed by each such group of concepts between the frequency of occurrences and the weighted cluster

for each such concept grouped into the group of concepts. Such claim is neither taught nor suggested by Marchisio and Holt.

In contrast, Claim 9 defines a method for analyzing unstructured documents for conceptual relationships. Claim 9 recites determining a frequency of occurrences of concepts in a set of unstructured documents, each concept representing an element occurring in one or more of the unstructured documents. Claim 9 further recites selecting a subset of concepts out of the frequency of occurrences. Claim 9 further recites grouping one or more concepts from the concepts subset. Claim 9 further recites assigning weights to one or more clusters of concepts for each group of concepts. Claim 9 further recites calculating a best fit approximation for each document indexed by each such group of concepts between the frequency of occurrences and the weighted cluster for each such concept grouped into the group of concepts. Such claim is neither taught nor suggested by Marchisio and Holt.

In contrast, Claim 18 defines a system for dynamically evaluating latent concepts in unstructured documents. Claim 18 recites an extraction module extracting a multiplicity of concepts from a set of unstructured documents into a lexicon uniquely identifying each concept and a frequency of occurrence. Claim 18 further recites a frequency mapping module creating a frequency of occurrence representation for each documents set, the representation providing an ordered corpus of the frequencies of occurrence of each concept. Claim 18 further recites a concept selection module selecting a subset of concepts from the frequency of occurrence representation filtered against a minimal set of concepts each referenced in at least two documents with no document in the corpus being unreferenced. Claim 18 further recites a group generation module generating a group of weighted clusters of concepts selected from the concepts subset. Claim 18 further recites a best fit module determining a matrix of best fit approximations for each document weighted against each group of weighted clusters of concepts. Such claim is neither taught nor suggested by Marchisio and Holt.

In contrast, Claim 31 defines a method for dynamically evaluating latent concepts in unstructured documents. Claim 31 recites extracting a multiplicity of

concepts from a set of unstructured documents into a lexicon uniquely identifying each concept and a frequency of occurrence. Claim 31 further recites creating a frequency of occurrence representation for each documents set, the representation providing an ordered corpus of the frequencies of occurrence of each concept. Claim 31 further recites selecting a subset of concepts from the frequency of occurrence representation filtered against a minimal set of concepts each referenced in at least two documents with no document in the corpus being unreferenced. Claim 31 further recites generating a group of weighted clusters of concepts selected from the concepts subset. Claim 31 further recites determining a matrix of best fit approximations for each document weighted against each group of weighted clusters of concepts. Such claim is neither taught nor suggested by Marchisio and Holt.

First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or combine the reference teachings and there must be a reasonable expectation of success. The teachings or suggestion to make the claimed combination and the reasonable expectation of success must both be found in the prior art, and not based on applicant's disclosure. MPEP § 2143 (citing *In re Vaeck*, 947 F.2d 488 (Fed. Cir. 1991)).

Marchisio fails to provide a suggestion or motivation to modify or combine with the reference teachings of Holt. Marchisio teaches formulating a constrained optimization problem in a *linear transform space* based on a term-spread matrix, error-covariance matrix and the user query vector, with the term-spread matrix corresponding to a weighted autocorrelation of the term-document matrix and indicating an amount of variation in term usage in the information files and extent to which the terms are correlated (Col. 22, lines 14-44). In contrast, Holt teaches utilizing *multidimensional subspaces* to represent semantic relationships that exist in a set of documents, wherein the documents are represented using a subspace transformation based on the distribution of the occurrence of terms in the documents (Col. 1, lines 19-23). In particular, a term-by-document frequency matrix is initially constructed that catalogs the

frequencies of the various terms for each of the documents, which is then preprocessed to define a working matrix by normalizing the columns of the term-by-document matrix to have a unit sum, stabilizing the variance of the term frequencies via a *non-linear* function and centering the term frequencies with respect to the mean vector of the columns (Col. 3, lines 16-28). Holt teaches away from the use of traditional vector space methods, such as used by Marchisio (*see*, Holt, Col. 2, lines 30-57), in favor of subspace transformations, by distinguishing the vector space methods as having performance severely limited by the size of the document collection, particularly with respect to recomputation of term weighting factors (Col. 2, line 64 through Col. 3, line 9; Col. 5, lines 30-40). As a result, one of ordinary skill in the art would not find a suggestion or motivation to combine the teachings of Marchisio with the teachings of Holt.

Similarly, one of ordinary skill in the art would not have a reasonable expectation of success in combining the teachings of Marchisio and Holt. The subspace transformation taught by Holt assigns weights to terms in a user query (Col. 5, lines 26-27), whereas Marchisio teaches an opposite approach of assigning weights to the elements of the term-document matrix (Col. 7, lines 10-17). Moreover, the traditional term weighting approach taught by Marchisio is rejected by Holt as requiring relatively time consuming and processing intensive recomputation upon the addition of new documents or removal of old documents from the document collection, as well as being unsuitable for some applications, such as the assignment of topic words, that is, words automatically generated to summarize a document (*see*, Holt, Col. 5, lines 32-46). Thus, combining the teachings of Marchisio and Holt would not result in a successful combination.

Finally, the combined references of Marchisio and Holt fail to teach or suggest all claim limitations. Preliminarily, Claims 1-17 define systems and methods for analyzing unstructured documents for conceptual relationships. Claims 18-44 define systems and methods for dynamically evaluating latent concepts in unstructured documents. Neither set of claims recites a user query that is analyzed against a collection of documents, as taught by the combination of Marchisio and Holt.

More particularly, Marchisio teaches adding a new *row* to the term-document matrix for each *phrase* in the user query, where each cell in the new row contains the frequency of occurrence of the phrase within the respective electronic information file, as determined by the frequencies of occurrence of individual terms composing the phrase and the proximity of such concepts, as determined by their relative positions in the electronic information files, as indicated by the elements of the auxiliary data structure. Thus, the term-document matrix grows, row-by-row, based on the phrases occurring in the user query, whereas the frequency of occurrences of concepts recited by Claims 1, 9, 18, and 31 are determined for the set of documents without reference to user query phrases, as taught by Marchisio. Nor does Marchisio suggest adding new term-document matrix rows independently from user query phrases.

Marchisio further teaches an auxiliary data structure that permits reforming of the term-document matrix to include rows corresponding to phrases in the user query for the purposes of processing that query. Thus, the rows in the auxiliary data structure are also dependent upon the phrases occurring in the user query, whereas the subsets of concepts recited by Claims 1, 9, 18, and 31 are selected out of the frequency of occurrences without reference to user query phrases, as taught by Marchisio. Nor does Marchisio suggest including the term-document matrix rows independently from user query phrases.

Marchisio further teaches assigning weights to the *elements* of the term-document matrix, whereas the weights recited by Claims 1, 9, 18, and 31 are assigned to one or more *clusters* of concepts for each group of concepts and not to a matrix, as taught by Marchisio. Nor does Marchisio suggest assigning weights to specifically formed clusters of concepts.

Finally, Holt teaches evaluating a score vector to determine the relative performance of the documents against the *user query*. The documents to return to a user are selected in a variety of methods, typically by returning the best scoring documents identified, for example, by applying a threshold to the individual scores, by taking a fixed number in ranked order, or by statistical or clustering techniques applied to the vectors of the scores. In contrast, the best fit

approximation recited by Claims 1, 9, 18, and 31 is calculated for each document indexed by each such group of concepts between the frequency of occurrences and the weighted cluster for each such concept grouped into the group of concepts and without reference to a user query, as taught by Holt. Nor does Holt suggest scoring documents independently from the user query. Thus, the combined references of Marchisio and Holt fail to teach or suggest all claim limitations.

Thus, a *prima facie* case of obviousness has not been shown with respect to Claims 1, 9, 18, and 31. Claims 2-8 are dependent on Claim 1 and are patentable for the above-stated reasons, and as further distinguished by the limitations recited therein. Claims 10-17 are dependent on Claim 9 and are patentable for the above-stated reasons, and as further distinguished by the limitations recited therein. Claims 19-30 are dependent on Claim 18 and are patentable for the above-stated reasons, and as further distinguished by the limitations recited therein. Claims 32-44 are dependent on Claim 31 and are patentable for the above-stated reasons, and as further distinguished by the limitations recited herein.

The prior art made of record and not relied upon has been reviewed by the applicant and is considered to be no more pertinent than the prior art references already applied.

Claims 1-44 are believed to be in a condition for allowance. Entry of the foregoing amendments is requested and a Notice of Allowance is earnestly solicited. Please contact the undersigned at (206) 381-3900 regarding any questions or concerns associated with the present matter.

Respectfully submitted,

Dated: August 27, 2004

By: ______________________
Patrick J.S. Inouye, Esq.
Reg. No. 40,297

Law Offices of Patrick J.S. Inouye
810 Third Avenue, Suite 258
Seattle, WA 98104

Telephone: (206) 381-3900
Facsimile: (206) 381-3999